BISS Biodiversity Information Science and Standards

Biodiversity Information Science and Standards 3: e37228
doi: 10.3897/biss.3.37228

OPEN ACCESS

Conference Abstract

LightningBug ONE: An experiment in high-throughput digitization of pinned insects

Mark Hereld‡, Nicola J. Ferrier‡

‡ Argonne National Laboratory, Lemont, Illinois, United States of America

Corresponding author: Mark Hereld (hereld@anl.gov)

Received: 12 Jun 2019 | Published: 19 Jun 2019

Citation: Hereld M, Ferrier N (2019) LightningBug ONE: An experiment in high-throughput digitization of pinned insects. Biodiversity Information Science and Standards 3: e37228. https://doi.org/10.3897/biss.3.37228

Abstract

Digital technology presents us with new and compelling opportunities for discovery when focused on the world's natural history collections. The outstanding barrier to applying existing and forthcoming computational methods for large-scale study of this important resource is that it is (largely) not yet in the digital realm. Without development of new and much faster methods for digitizing objects in these collections, it will be a long time before these data are available in digital form. For example, methods that are currently employed for capturing, cataloguing, and indexing pinned insect specimen data will require many tens of years or more to process collections with millions of dry specimens, and so we need to develop a much faster pipeline. In this paper we describe a capture system capable of collecting and archiving the imagery necessary to digitize a collection of circa 4.5 million specimens in one or two years of production operation.

To minimize the time required to digitize each specimen, we have proposed (Hereld et al. 2017) developing multi-camera systems to capture the pinned insect and its accompanying labels from many angles in a single exposure. Using a sampling (21 randomly drawn drawers, totalling 5178 insects) of the 4.5 million specimens in the collection at the Field Museum of Natural History, we estimated that a large fraction of that collection (97.6% +/- 2.2%) consists of pinned insects with labels that are visible from one angle or another without requiring adjustment or removal of elements on the pin. In this situation a multi-camera system with enough angular coverage could provide imagery for reconstructing virtual labels from fragmentary views taken from different directions. Agarwal et al.

© Hereld M, Ferrier N. This is an open access article distributed under the terms of the Creative Commons Attribution License (CC BY 4.0), which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited.

(2018) demonstrated a method for combining these multiple views into a virtual label that could be transcribed by automated optical character recognition software.

We have now designed, built and tested a prototype snapshot 3D digitization station to allow rapid capture of multi-view imagery for automated capture of pinned insect specimens and labels. It consists of twelve very small and light 8-megapixel cameras (Fig. 1), each controlled by a small dedicated computer. The cameras are arrayed around the target volume, six on each side of the sample feed path. Their positions and orientations are fixed by a 3D-printed scaffolding designed for the purpose. The twelve camera controllers and a master computer are connected to a dedicated high-speed data network over which all of the coordinating control signals and returning images and metadata are passed. The system is integrated with a high-performance object store that includes a database for metadata and the archived images comprising each snapshot. The system is designed so that it can be readily extended to include additional or different sensors.

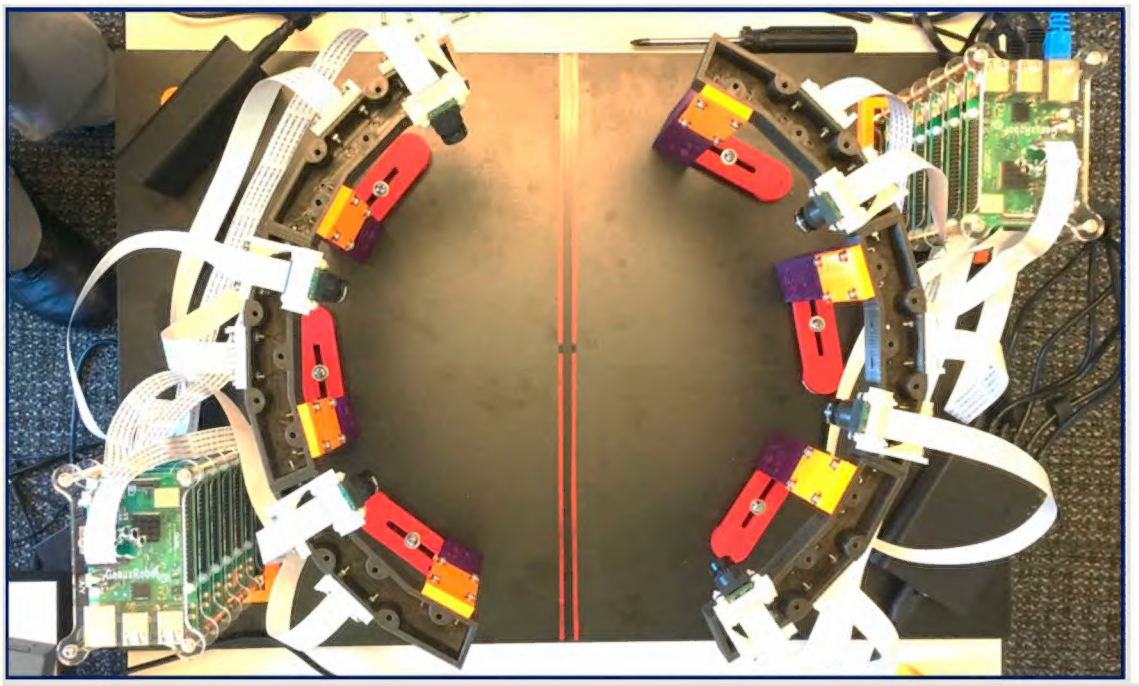

Figure 1.

The prototype snapshot 3D digitization station as viewed from above. The specimen feed path is indicated by the vertical stripe with red trim at the center of the image. For scale, the marked path is 16 inches long. Scaffolding, printed in several colors of PLA plastic, to the left and right of this path provide for fixed placement of six cameras each. Three are visibly mounted above the semi-circular rail and three are less-visibly mounted below. Each is connected via ribbon cable to its control computer in the correpsonding stack at the lower left and upper right of the image.

The station is meant to be fed with specimens by a conveyor belt whose motion is coordinated with the exposure of the multi-view snapshots. In order to test the performance of the system we added a recirculating specimen feeder designed expressly for this experiment. With it integrated into the system in place of a conventional conveyor belt we are able to provide a continuous stream of targets for the digitization system to facilitate long tests of its performance and robustness. We demonstrated the ability to capture data at a peak rate of 1400 specimens per hour and an average rate of 1000 specimens per

hour over the course of a sustained 6 hour run. The dataset (Hereld and Ferrier 2018) collected in this experiment provides fodder for the further development of algorithms for the offline reconstruction and automatic transcription of the label contents.

Keywords

mass digitization, pinned insects, image analysis, multiple view snapshot imaging

Presenting author

Mark Hereld

Acknowledgements

We thank Petra Sierwald, Rüdiger Bieler, and Crystal Maier for many informative discussions about workflows and current practices. This material is based upon work supported by Laboratory Directed Research and Development (LDRD) funding from Argonne National Laboratory, provided by the Director, Office of Science, of the U.S. Department of Energy under contract DE-AC02-06CH11357. Additional support was provided by the Field Museum of Natural History.

References

- Agarwal N, Ferrier N, Hereld M (2018) Towards Automated Transcription of Label Text from Pinned Insect Collections. 2018 IEEE Winter Conference on Applications of Computer Vision (WACV) https://doi.org/10.1109/wacv.2018.00027
- Hereld M, Ferrier N, Agarwal N, Sierwald P (2017) Designing a High-Throughput Pipeline for Digitizing Pinned Insects. 2017 IEEE 13th International Conference on e-Science (e-Science) https://doi.org/10.1109/escience.2017.88
- Hereld M, Ferrier N (2018) LightningBug ONE, a multi-view snapshot dataset for development of automated digitization algorithms. https://silo18.github.io/LightningBugONE/. Accessed on: 2019-4-08.